26º ENANCIB
Encontro Nacional de Pesquisa e Pós-Graduação em Ciência da Informação
A Amazônia na sociedade global em rede: assimetrias e potencialidades informacionais, tecnológicas e ambientais
Belém, Pará, 09 a 13 de novembro de 2026
Universidade Federal do Pará

ENCONTRO NACIONAL DE PESQUISA E PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO

GT 8 – Dados, Informação e Tecnologia

INTEROPERABILIDADE DE METADADOS DE REPOSITÓRIOS INSTITUCIONAIS PARA A PLATAFORMA SUCUPIRA: PADRÃO OAI-CAPES

METADATA INTEROPERABILITY FROM INSTITUTIONAL REPOSITORIES TO THE SUCUPIRA PLATFORM: THE OAI-CAPES STANDARD

Caio Saraiva Coneglian – Universidade de Marília (UNIMAR), Universidade Estadual Paulista (UNESP) – Campus Marília e Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT)
Emanuelle Torino – Universidade Estadual Paulista (UNESP) – Campus Marília, Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT) e Universidade Tecnológica Federal do Paraná (UTFPR)
Henrique Leal Tavares – Universidade de Marília (UNIMAR) e Universidade Estadual Paulista (UNESP) – Campus Marília
Mariana Baptista Brandt - Câmara dos Deputados
Silvana Aparecida Borsetti Gregorio Vidotti - Universidade Estadual Paulista (UNESP) – Campus Marília e Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT)

Modalidade: Trabalho Completo

Resumo: Os repositórios institucionais se consolidaram como um dos principais ambientes de disseminação e de gestão das produções intelectuais das instituições, mostrando-se como uma importante fonte para a coleta de dados sobre teses e dissertações defendidas no contexto nacional. No entanto, diante da grande heterogeneidade dos perfis de aplicação existentes nos repositórios digitais, surge a necessidade de criar meios de interoperabilidade e de padronização dos registros quando se busca obtê-los para utilizá-los como fontes informacionais para alimentar o processo de avaliação da pós-graduação brasileira. Dessa forma, este trabalho objetiva apresentar os aspectos teórico-práticos utilizados para converter dados de repositórios institucionais para o formato utilizado pela Plataforma Sucupira, possibilitando operacionalizar a integração automática desses dados. Como procedimento metodológico, adotou-se uma pesquisa aplicada de natureza qualitativa, fundamentada em análise documental das especificações técnicas da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES). Como resultados, tem-se o desenvolvimento de três entregáveis que estão disponíveis para a comunidade acadêmica: o

Guia de Instalação do oai-capes, o arquivo XSLT padrão para a realização do *crosswalk* dos registros e a planilha de conversão para utilização pelos gestores de repositórios. Aponta-se, ainda, que o padrão já foi validado em dois contextos institucionais. Portanto, conclui-se que o padrão oai-capes constitui uma alternativa viável para a coleta de dados pela Plataforma Sucupira de Teses e Dissertações, posicionando os repositórios institucionais como fontes estratégicas para o processo de avaliação da pós-graduação.
Palavras-chave: repositório institucional; metadados; integração de dados; Protocolo OAI-PMH; padrão oai-capes.

Abstract: Institutional repositories have become established as key environments for the dissemination and management of academic intellectual output, emerging as important sources for collecting data on theses and dissertations defended in Brazil. However, given the marked heterogeneity of the application profiles found across digital repositories, there is a need to establish mechanisms for interoperability and record standardization when data are retrieved and used as information sources to support the Brazilian postgraduate education evaluation process. Accordingly, this work aims to present the theoretical and practical aspects involved in converting data from institutional repositories into the format used by the Sucupira Platform, making it possible to implement the automatic integration of these data. An applied, qualitative research approach was adopted, based on a documentary analysis of the technical specifications of the Coordination for the Improvement of Higher Education Personnel (CAPES). As a result, three deliverables were developed and are available to the academic community: the oai-capes Installation Guide, the standard XSLT file for performing the record crosswalk, and the conversion spreadsheet for use by repository managers. It is further noted that the standard has already been validated in two institutional contexts. Therefore, the study concludes that the oai-capes standard constitutes a viable alternative for the collection of thesis and dissertation data by the Sucupira Platform, positioning institutional repositories as strategic sources for the postgraduate education evaluation process.
Keywords: institutional repository; metadata; data integration; OAI-PMH Protocol; oai-capes standard.

1 INTRODUÇÃO

A Avaliação dos Programas de Pós-Graduação no Brasil, estabelecida em 1998 como Avaliação do Sistema Nacional de Pós-Graduação, visa assegurar a qualidade dos cursos de pós-graduação *stricto sensu* brasileiros. Ela é conduzida com a participação da comunidade, por meio de consultores *ad hoc*, sob orientação da Diretoria de Avaliação (DAV) da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES).

Esse processo avaliativo se realiza com base nos documentos de área e nas fichas de avaliação, a partir da análise de dados informados pelos programas de pós-graduação na Plataforma Sucupira. Cumpre destacar que cada uma das 50 áreas de avaliação[1] possui um

[1] Disponível em: https://www.gov.br/capes/pt-br/acesso-a-informacao/acoes-e-programas/avaliacao/sobre-a-avaliacao/areas-avaliacao/sobre-as-areas-de-avaliacao/sobre-as-areas-de-avaliacao.

documento de área específico que estabelece quesitos prioritários na avaliação.

No ano de 2023, a CAPES lançou o Programa de Governança Colaborativa de Informações da Pós-Graduação (GoPG), por meio da Portaria CAPES nº 158, de 17 de agosto de 2023, visando aprimorar o processo de coleta dos dados e reduzir o trabalho operacional (Brasil, 2023).

Para tanto, a CAPES tem disponibilizado às Instituições de Ensino Superior e de Pesquisa (IES) e aos Programas de Pós-Graduação (PPG) um conjunto de ferramentas, padrões, processos, diretrizes, guias e boas práticas, visando auxiliar na implementação prática do Programa GoPG pelas IES interessadas em fornecer os dados para o processo de avaliação da pós-graduação de forma automatizada. As informações sobre todo esse aparato estão disponíveis no Guia de Arquitetura e Interoperabilidade de Dados da Pós-Graduação[2], que está em constante processo de ampliação e aperfeiçoamento.

Dentre os dados que fazem parte do processo de avaliação estão as produções intelectuais, que podem ser de natureza bibliográfica, técnica e tecnológica, artística ou cultural. Os dados de produção intelectual podem ser enviados pelos PPG para a Plataforma Sucupira de forma automatizada, utilizando o que preconiza o GoPG. No contexto das produções intelectuais bibliográficas, as dissertações e teses são tipologias que há muitos anos são gerenciadas em cenário nacional pelas bibliotecas que mais recentemente vêm utilizando os Repositórios Institucionais (RI) como ambientes informacionais.

Um projeto piloto realizado pela CAPES (utilizando, inicialmente, quatro repositórios brasileiros e depois expandido para 18, considerando todas as IES que em 2025 haviam aderido ao GoPG e que possuem DSpace para o gerenciamento de dissertações e teses) apontou que a totalidade dos RI está com a saída OAI-PMH configurada e exibe os dados em xoai[3].

Um processo inicial de extração de dados desses 18 repositórios, via OAI-PMH e utilizando xoai, indicou que, embora haja correspondência com muitos metadados requeridos pela CAPES para representar dissertações e teses, é grande a variedade de metadados utilizados nas estruturas, o que tornaria necessário um processo de análise e compatibilização, o qual se mostra inviável ao considerarmos que há mais de 500 IES no Sistema Nacional de

[2] Disponível em: https://sucupira.capes.gov.br/guia-de-interoperabilidade/introducao/guia-pratico.
[3] O OAI-PMH não fornece diretrizes para saídas de dados específicas; assim, o xoai permite transformar os metadados, de forma a atender a requisitos específicos por meio de configuração.

Pós-Graduação (SNPG).

Diante do exposto, o problema de pesquisa consiste em identificar uma alternativa viável para a conversão de metadados de RI que utilizam perfis de aplicação próprios para a estrutura de dados da Plataforma Sucupira, reduzindo os custos operacionais, os ruídos, a inconsistência, a obsolescência e a possibilidade de falhas.

Assim, o presente trabalho objetiva apresentar os aspectos teórico-práticos utilizados para converter dados de repositórios institucionais para o formato utilizado pela Plataforma Sucupira, possibilitando operacionalizar a integração automática de dados estabelecida pelo GoPG.

O processo culminou no desenvolvimento do Padrão oai-capes, uma saída de dados específica que utiliza o xoai para transformar os metadados locais de uma instância de DSpace para o Perfil de Aplicação da CAPES, o que requereu o desenvolvimento de um *crosswalk* de metadados no formato *Extensible Stylesheet Language Transformations* (XSLT) que realiza o mapeamento dos metadados e os expõe no padrão CAPES. Atualmente, o Padrão oai-capes encontra-se na versão 1.2, está estruturado, inicialmente, para dissertações e teses e será expandido à medida que outros tipos de produção intelectual forem requeridos pela CAPES. O Padrão oai-capes foi documentado, está disponibilizado publicamente[4] e seu processo é explicitado neste texto.

A relevância desta pesquisa reside tanto no aspecto prático da integração de dados entre repositórios institucionais e sistemas nacionais de coleta de dados quanto no aspecto teórico, ao evidenciar como a conversão de registros pode ser estruturada visando possibilitar que infraestruturas heterogêneas tenham seus dados integrados.

2 REFERENCIAL TEÓRICO

A Rede de Integração da Comunidade Acadêmico-Científica da Pós-Graduação (RICA|PG)[5] foi constituída no ano de 2020 e, até o ano de 2022, na chamada primeira fase, atuou em aspectos relacionados à gestão da rede e à compreensão da forma de armazenamento dos dados da pós-graduação nas IES, bem como dos sistemas de informação

[4] Disponível em: https://sucupira.capes.gov.br/guia-de-interoperabilidade/etapas-implementacao/etapa3-interoperabilidade-tecnica/protocolo-repositorios/oai-pmh-capes.
[5] Disponível em: https://www.gov.br/capes/pt-br/acesso-a-informacao/acoes-e-programas/avaliacao/programa-de-governanca-colaborativa-de-informacoes-da-pos-graduacao-gopg/rede-rica.

utilizados e de seus níveis de maturidade. Na primeira fase, considerando o cenário das instituições mapeadas, foi elaborada a proposta de um conjunto de dados mínimos que pudessem ser coletados de todas as IES.

No ano de 2023, a CAPES lançou o Programa de Governança Colaborativa de Informações da Pós-Graduação (GoPG)[6], que atribuiu à Rede de Integração da Comunidade Acadêmico-Científica da Pós-Graduação (RICA|PG) o estabelecimento de "Padrões a serem adotados para troca de dados"[7] no âmbito da avaliação da pós-graduação brasileira.

Nesse sentido, em outubro de 2023, por demanda da Diretoria de Avaliação (DAV) da CAPES, as atividades da RICA|PG foram retomadas, em uma segunda fase, estruturada em um projeto de desenvolvimento científico e tecnológico intitulado "Arquitetura de Dados da Avaliação da Pós-Graduação Brasileira no Contexto da CAPES". Entre os processos realizados no projeto, está a definição de uma estrutura de metadados a ser utilizada pela Plataforma Sucupira, desenvolvida a partir de um Perfil de Aplicação Ampliado, que agrega questões descritivas, técnicas, de processo de trabalho, de fluxo, de governança, entre outras.

A modelagem está sendo desenvolvida considerando entidades e seus relacionamentos, dos quais destacam-se, neste trabalho, a Entidade Produção Intelectual do tipo Bibliográfica e a entidade especializada Trabalho de Conclusão de Curso (TCC), que abrange dissertações e teses.

2.1 INTEROPERABILIDADE, METADADOS E CROSSWALK DE METADADOS

A interoperabilidade designa a capacidade de dois ou mais sistemas trocarem e utilizarem dados sem perda de significado e sem intervenção manual. No campo dos metadados, distinguem-se três níveis complementares (Chan; Zeng, 2006; Haslhofer; Klas, 2010). A interoperabilidade sintática refere-se à compatibilidade de formatos e estruturas de serialização. A semântica diz respeito à equivalência de significado entre elementos e valores: não basta que o registro seja legível, é preciso que o elemento signifique o mesmo em ambos os sistemas. A organizacional envolve acordos, políticas, responsabilidades e governança que sustentam a troca ao longo do tempo.

Dessa forma, a interoperabilidade realizada no contexto deste projeto ocorre no

[6] Disponível em: https://www.gov.br/capes/pt-br/acesso-a-informacao/acoes-e-programas/avaliacao/programa-de-governanca-colaborativa-de-informacoes-da-pos-graduacao-gopg.
[7] Disponível em: https://sucupira.capes.gov.br/guia-de-interoperabilidade/introducao/programa-governanca-gopg/rede-integracao-rica-pg.

âmbito dos metadados. Os metadados são "[...] estruturas de organização da informação, legíveis por máquinas, que visam tornar os dados úteis, de diferentes formas, de acordo com as necessidades específicas de cada serviço de informação digital e de acordo com a aplicação que lhes é concedida" (Méndez Rodríguez, 2002, p. 47, tradução nossa).

A adoção de padrões de metadados internacionalmente reconhecidos e especificados é o que possibilita o intercâmbio de dados e registros entre sistemas de informação, reduzindo o ruído de comunicação. Destaca-se que há uma ampla variedade de padrões de metadados disponíveis e que sua utilização requer conhecimento do objetivo e do domínio de uso, justamente por cada padrão possuir especificidades, diretrizes e extensão.

Em aplicações específicas, o uso de padrões de metadados é especificado por um perfil de aplicação de metadados. "Definimos perfis de aplicação como esquemas que consistem em elementos de dados extraídos de um ou mais padrões de metadados, combinados pelos implementadores e otimizados para uma aplicação local específica" (Heery; Patel, 2000). Tem-se, assim, a possibilidade de uso de um ou mais padrões de metadados em um ambiente informacional digital, e a função do perfil de aplicação é registrar e estabelecer as formas de uso de cada elemento. Para Heery e Patel (2000), o perfil de aplicação permite ao implementador expressar como está utilizando os esquemas.

Dessa forma, com base na literatura e considerando as especificidades e contextos práticos de implementação de metadados, é imprescindível que o perfil de aplicação seja desenvolvido com alinhamento ao domínio, ao contexto e à aplicação, o que torna cada perfil de aplicação único. Vale destacar que o Singapore Framework (Nilsson; Baker; Johnston, 2008) sistematiza os componentes que um perfil deve documentar: requisitos funcionais, modelo de domínio, descrição dos termos, diretrizes de uso e sintaxes admitidas.

No processo de interoperabilidade de metadados, destaca-se que o *crosswalk* é o instrumento mais difundido para promover a troca entre esquemas distintos, especificando as equivalências entre os elementos de um esquema de origem e os de um esquema de destino, com as regras de transformação correspondentes (St. Pierre; LaPlant, 1998; Woodley, 2016). Essas equivalências assumem cardinalidades diversas: um-para-um, quando há correspondência unívoca; um-para-muitos, quando um elemento precisa ser decomposto, como no campo que reúne nome e identificador de uma pessoa; e muitos-para-um, quando vários elementos são agregados em um só, caso em que a granularidade original deixa de ser recuperável. Toda transformação comporta algum grau de perda semântica e traz, assim, a

supressão de distinções não representáveis no esquema de destino. Daí a distinção entre transformações reversíveis e não reversíveis: o *crosswalk* aqui proposto é não reversível, pois preserva intacto o perfil local, mas sua saída não permite reconstituir integralmente o registro de origem.

2.2 PROTOCOLO OAI-PMH

O *Open Archives Initiative Protocol for Metadata Harvesting* (OAI-PMH) é um mecanismo desenvolvido para possibilitar o fornecimento e o consumo de dados entre provedores de dados e de serviços. Nesse contexto, os provedores de dados expõem metadados estruturados em uma saída OAI-PMH para a qual podem ser enviadas requisições de coleta. As requisições são feitas via *Hypertext Transfer Protocol* (HTTP), e o OAI-PMH responde em um conjunto de seis verbos: *Identify, ListMetadataFormats, ListSets, ListIdentifiers, ListRecords e GetRecord*. Todas as respostas a solicitações feitas pelo OAI-PMH são documentos em *Extensible Markup Language*, linguagem representada pela sigla XML (Open Archives Initiative, 2002).

No contexto do provedor de dados, o OAI-PMH considera três elementos: recurso, que consiste na entidade ou no objeto ao qual os metadados se referem; item, componente que possibilita a disseminação de um recurso; e registro, que consiste nos metadados expressos em um formato específico (Open Archives Initiative, 2002).

O protocolo OAI-PMH suporta o fornecimento de metadados em múltiplos formatos, que podem ser consultados a partir de uma requisição *ListMetadataFormats*. Cada formato possui as seguintes propriedades: *metadataPrefix*, os caracteres válidos utilizados para exibir ou recuperar registros nas requisições *ListRecords*, *ListIdentifiers* e *GetRecord*, cujo retorno se dará no formato especificado; *metadata schema* URL, que consiste na URL que possibilita verificar os metadados de acordo com o formato; e XML *namespace*, URI que identifica o formato de metadados (Open Archives Initiative, 2002).

Nesse contexto, usando o DSpace como exemplo de software para repositórios institucionais, a documentação especifica que há uma saída designada xoai[8] que permite criar saídas mais flexíveis para o fornecimento de dados via OAI-PMH.

Destaca-se que o OAI-PMH, apesar de sua ampla difusão, apresenta limitações

[8] Disponível em: https://github.com/DSpace/xoai.

reconhecidas. Tal protocolo foi projetado para a coleta periódica de metadados descritivos em lote, e não para acesso granular ou em tempo real (Van de Sompel; Nelson, 2000; Van de Sompel; Nelson, 2015). A escolha se justifica, ainda assim, pela base instalada, já que a totalidade dos repositórios analisados já expõe dados por esse protocolo, pela maturidade das ferramentas no DSpace e pela possibilidade de adotar formatos adicionais, o que permite expor um perfil específico sem alterar o perfil local. Alternativas como APIs REST ou *ResourceSync* implicariam esforço de desenvolvimento incompatível com a escala e a heterogeneidade tecnológica do SNPG.

3 PROCEDIMENTOS METODOLÓGICOS

Esta pesquisa caracteriza-se como aplicada, de natureza qualitativa, fundamentada em análise documental e em desenvolvimento tecnológico. O percurso metodológico que culminou no desenvolvimento do padrão oai-capes abrangeu cinco etapas complementares:

a) análise dos metadados utilizados por RI das IES que aderiram ao GoPG até o ano de 2025, cujo resultado apontou heterogeneidade entre as estruturas de dados dos RI, o que amplia a problemática, considerando o Perfil de Aplicação da própria Plataforma Sucupira. No total, foram analisados 960 campos de metadados dos repositórios estudados. Esse processo levou a uma modelagem conceitual do padrão oai-capes;
b) levantamento e análise das especificações técnicas do DSpace quanto à saída de dados via OAI-PMH, incluindo aquelas relativas ao xoai e à configuração do protocolo OAI-PMH. Paralelamente, foram analisadas as especificações técnicas do próprio OAI-PMH. Esse processo embasou a proposição da implementação técnica do padrão oai-capes, por meio do *crosswalk* de metadados;
c) coleta de dados via xoai em 18 repositórios institucionais vinculados a 21 IES que aderiram ao GoPG até o ano de 2025, visando compreender as estruturas de dados e os próprios dados. Essa etapa, ocorrida agosto de 2025, possibilitou planejar e propor a estrutura do *crosswalk* de metadados, utilizando uma planilha-base para a compatibilização entre a estrutura de dados dos RI e o Perfil de Aplicação da Plataforma Sucupira;
d) desenvolvimento iterativo do *crosswalk* de metadados, buscando validar a estrutura de dois documentos: a planilha para o mapeamento dos metadados entre os RI e a Plataforma Sucupira e o arquivo *Extensible Stylesheet Language Transformations* (XSLT)

que materializa o *crosswalk* de metadados e possibilitou a estruturação de um Guia de Configuração do oai-capes;

e) aplicação e validação do *crosswalk* de metadados em contextos reais, de forma iterativa entre desenvolvedores e IES, por meio do preenchimento da planilha de mapeamento de metadados e de seu uso posterior para adequação do modelo de XSLT para o *crosswalk* de metadados. Esse processo possibilitou validar os documentos gerados: o Guia de Configuração do oai-capes, a planilha para o mapeamento dos metadados entre os RI e a Plataforma Sucupira e o arquivo *Extensible Stylesheet Language Transformations* (XSLT) que materializa o *crosswalk* de metadados. Essa documentação foi disponibilizada publicamente pela CAPES para utilização pelas IES interessadas.

Esse percurso metodológico possibilitou, com base na pesquisa documental, desenvolver uma aplicação tecnológica, disponibilizada publicamente para a comunidade.

4 RESULTADOS: O PADRÃO OAI-CAPES

A CAPES publicou a Instrução Normativa nº 1, de 23 de abril de 2026, que "Dispõe sobre a operacionalização da interoperabilidade e do compartilhamento de dados no âmbito do Programa de Governança Colaborativa de Informações da Pós-Graduação (GoPG)" e estabelece, no artigo 13, as fontes originárias de interoperabilidade, indicando os repositórios institucionais como "responsáveis pela produção intelectual, compreendendo dissertações, teses, artigos, livros, capítulos, trabalhos em eventos vinculados aos PPGs" (Brasil, 2026, não paginado). A referida Instrução Normativa estabelece ainda que "em cada uma dessas fontes haverá conjuntos de dados específicos e serviços diferenciados de interoperabilidade, conforme padrões técnicos e semânticos definidos pela CAPES."

Para a operacionalização desse processo de integração de dados de repositórios institucionais, são apresentados neste trabalho os processos estabelecidos metodologicamente para o desenvolvimento do padrão oai-capes, que possibilita que os dados dos RI sejam convertidos para o padrão de dados da Plataforma Sucupira da CAPES.

4.1 COLETA E ANÁLISE DE DADOS DOS REPOSITÓRIOS INSTITUCIONAIS

Visando analisar a estrutura de dados de teses e dissertações em repositórios institucionais brasileiros, foi conduzido um projeto piloto de extração de dados, via OAI-PMH, em duas fases. A primeira fase envolveu quatro RI e tinha como objetivo validar a

correspondência entre os metadados requeridos pela CAPES e aqueles disponibilizados nos RI. Foi identificada correspondência entre os metadados, porém havia divergências na estrutura semântica dos metadados, além de alguns metadados faltantes.

A segunda fase expandiu a análise para as 21 IES que haviam aderido ao GoPG até o ano de 2025. Entre elas, 18 disponibilizam RI, dos quais foram coletados, via OAI-PMH, os metadados de teses e dissertações defendidas nos anos de 2025 e 2026. Após a coleta, foi possível identificar que muitos metadados requeridos pela CAPES para representar dissertações e teses estão disponíveis nos RI, embora haja uma grande heterogeneidade na estrutura de representação e divergências significativas na estrutura semântica, além do uso de diferentes termos semânticos para o mesmo metadado.

Esse piloto evidenciou que o uso de dados de RI para alimentar a Plataforma Sucupira necessitaria de um processo de compatibilização, considerando que cada RI possui especificidades em sua infraestrutura. O processo poderia ser executado, porém a complexidade de realização e da manutenção de estruturas de conversão de registros, diretamente no processamento realizado pela CAPES, torna-o inviável (considerando que o SNPG é composto por aproximadamente 500 IES), sobretudo ao considerar a possibilidade de interpretações imprecisas na compatibilização, alterações que levariam à obsolescência, custo operacional de manutenção, ruídos e falhas na comunicação de dados.

Assim, diante da falta de um padrão no processo de representação, e considerando que a CAPES reconhece a autonomia universitária, a proposta viável para que os dados dos RI possam ser utilizados de forma automática para o preenchimento da Plataforma Sucupira, sem incorrer nos problemas expostos anteriormente, foi estabelecer uma estrutura de conversão de metadados implementada diretamente no RI, que passa a expor os dados, via OAI-PMH, no padrão especificado pela CAPES, o chamado Padrão oai-capes.

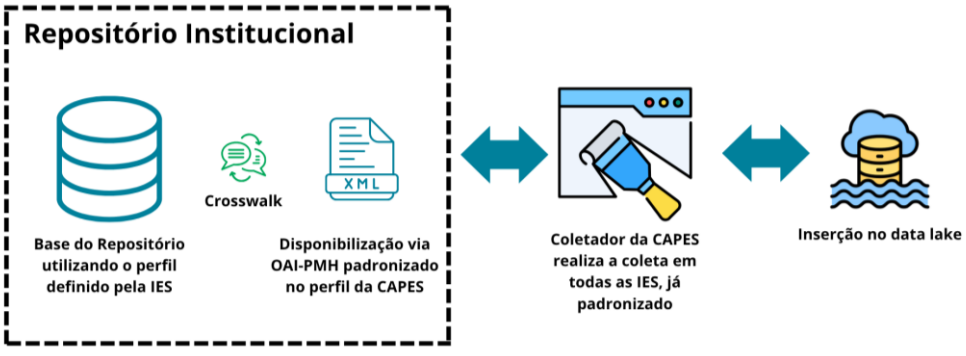
4.2 PROPOSTA DO PADRÃO OAI-CAPES

A criação do oai-capes, enquanto um padrão de fornecimento de dados via OAI-PMH, consiste em uma alternativa viável para que os dados armazenados em repositórios institucionais possam ser convertidos de forma padronizada para a integração com a Plataforma Sucupira da CAPES. Dessa forma, os repositórios institucionais mantêm seu perfil de aplicação inalterado e realizam uma conversão para o formato de metadados esperado pela CAPES, por meio de um *crosswalk* de metadados.

Cabe ressaltar dois elementos que embasam a criação do padrão oai-capes. O primeiro é relativo à adoção de vocabulários semânticos internacionais. O padrão oai-capes expõe os metadados utilizando termos semânticos de vocabulários internacionais e termos semânticos próprios do vocabulário CAPES, quando não existia termo internacional com o mesmo significado. Tais termos foram definidos no processo de arquitetura de dados, por meio da criação do Perfil de Aplicação Ampliado para a Pós-Graduação brasileira. A escolha por converter os metadados e expô-los utilizando termos semânticos reconhecidos internacionalmente justifica-se pela possibilidade de interoperar tais registros com outros domínios, além de permitir a compreensão do significado desses termos por humanos e agentes ou aplicações computacionais. O segundo elemento diz respeito à utilização do OAI-PMH. Esse protocolo foi escolhido como base para a requisição e consumo de registros, por ser difundido na comunidade acadêmica e profissional de repositórios digitais no contexto de integração de metadados. Há diversas aplicações que coletam os registros dos repositórios por meio do OAI-PMH, e o software DSpace, amplamente utilizado no Brasil para a implantação de repositórios, já oferece nativamente suporte ao OAI-PMH.

Partindo dos aspectos apontados, a Figura 1 apresenta o modelo do Padrão oai-capes, demonstrando a integração dos repositórios institucionais no âmbito das coletas de dados realizadas pela CAPES para o processo de avaliação.

Figura 1 - Modelo do Padrão oai-capes



Fonte: Elaborado pelos autores (2026).

A Figura 1 demonstra que, primeiramente, o repositório institucional mantido pelas IES já está estruturado com um perfil de aplicação de metadados próprio, sobre o qual se aplica um procedimento de *crosswalk*, que irá converter os metadados dos RI para o padrão adotado pela CAPES. Esses metadados serão exibidos, via protocolo OAI-PMH, em um

metadataPrefix designado “oai_capes”, em formato XML. Na segunda camada, o coletor da CAPES realiza a coleta automatizada e padronizada individualmente em cada RI, para consumir os dados normalizados, o que assegura consistência entre as fontes, caso o processo seja realizado corretamente. Por fim, os metadados coletados são inseridos em um *data lake*, um ambiente de armazenamento centralizado que consolida todos os dados necessários para o processo de avaliação da pós-graduação brasileira, instalado na CAPES e de onde os dados são consumidos pela Plataforma Sucupira. Dessa forma, a proposta do padrão oai-capes apresenta uma arquitetura de integração de dados baseada em padronização, na qual a heterogeneidade dos perfis de aplicação é tratada por meio do *crosswalk* de metadados, que favorece a escalabilidade e a integridade semântica na agregação de diversas fontes.

Cabe registrar, contudo, que a padronização opera sobre a estrutura e a terminologia da saída. Assim, não corrige metadados ausentes, valores inconsistentes ou equivalências inadequadas na origem, pois a qualidade dos metadados permanece condicionada às práticas descritivas de cada instituição.

4.2.1 Mapeamento de metadados do padrão oai-capes

O processo de mapeamento dos metadados para a conversão do perfil de aplicação do RI para o perfil de aplicação da CAPES é uma das bases para o *crosswalk* de metadados, fundamental para o padrão oai-capes.

Para tanto, foi desenvolvida e disponibilizada uma planilha para o mapeamento dos metadados entre os RI e a Plataforma Sucupira, que serve de guia para a compatibilização das estruturas. Essa planilha pode ser preenchida pela equipe do repositório institucional, tendo como base o Perfil de Aplicação do RI, cujos metadados semânticos podem, inclusive, ser identificados na interface destinada aos usuários, no caso do DSpace.

O Quadro 1 apresenta um trecho da referida planilha, que, em sua versão completa, é composta por outras colunas além das apresentadas neste fragmento.

Quadro 1 - Fragmento da planilha para o mapeamento dos metadados entre RI e Plataforma Sucupira

Metadado no Repositório da IES Indicar o termo utilizado no	Perfil de Aplicação da CAPES para Tese/Dissertação		
	Nome do Metadado	Termo semântico	Entrada Padronizad

repositório da IES para exibir o metadado correspondente			a
dc.type*	Tipo de documento da produção intelectual	dcterms:type	Lista do COAR
dc.creator*	Autor da produção intelectual	dcterms:creator	Não se aplica
dc.identifier.orcid	Identificador ORCiD da pessoa da produção intelectual (Autor)	vivo:orcidId	Não se aplica

Fonte: Elaborado pelos autores (2026).
*Os dados apresentados foram inseridos em caráter ilustrativo, para exemplificar a forma de preenchimento. Na planilha disponibilizada para uso dos RI, a primeira coluna está vazia e deve ser preenchida considerando o Perfil de Aplicação.

No momento do uso da planilha para o mapeamento dos metadados entre os RI e a Plataforma Sucupira, a primeira coluna deverá ser preenchida de acordo com o Perfil de Aplicação do Repositório Institucional.

Vale destacar que, no Perfil de Aplicação da CAPES, cujos campos utilizados para dissertações e teses estão disponíveis na planilha, foram utilizados termos semânticos de diversos padrões de metadados. Contudo, considerando as especificidades do contexto brasileiro e que, em alguns casos, não foram identificados termos semânticos cuja aplicação fosse possível, foi criado o vocabulário CAPES, que está em processo de especificação e disponibilização.

Após o processo de mapeamento, a planilha preenchida será utilizada como base para a elaboração de um arquivo XSLT para efetivamente realizar o *crosswalk* de metadados e disponibilizar de forma pública os metadados no padrão oai-capes.

4.2.2 Estrutura e padrões do crosswalk XSLT

O arquivo *Extensible Stylesheet Language Transformations* (XSLT) disponibilizado como modelo (*oai_capes.xsl*) consiste em uma folha de estilo XSLT, que se estrutura em torno de dois padrões principais de projeto: o template emit-multi e o uso de elementos contêineres sem prefixo de *namespace*. Destaca-se que alguns dos pontos apresentados nesta subseção fazem referência ao DSpace, considerando que se trata de um software *open source*

amplamente utilizado e que apresenta documentação extensa acerca da criação de padrões OAI-PMH.

O padrão emit-multi foi adotado para o tratamento de campos multivalorados. No modelo xoai do DSpace, campos como título, autor e membros de banca podem ocorrer múltiplas vezes como nós irmãos na árvore XML. O uso de *xsl:value-of* nesses casos capturaria apenas o primeiro valor; a iteração via *xsl:for-each* é o mecanismo correto para garantir que todos os valores sejam emitidos. A criação de um template nomeado reutilizável para esse padrão reduz a repetição de código e facilita a manutenção da folha de estilo.

O segundo padrão diz respeito ao uso correto de prefixos de *namespace* nos elementos de saída. Uma falha comum em *crosswalks* XSLT é declarar os *namespaces* no elemento raiz da folha de estilo sem incluí-los explicitamente nos nomes dos elementos de saída. O processador XSLT 1.0 do DSpace exige que os prefixos (como *dcterms:*, *capes:*, *vivo:* e *schema:*) apareçam nos nomes dos elementos de saída (por exemplo, *<dcterms:issued>*) e não apenas nas declarações de *namespace*. Em contraste, os elementos contêineres estruturais (como *<creator>*, *<advisor>* e *<committeeMember>*) não carregam prefixo de *namespace*, pois funcionam apenas como agrupadores estruturais dos dados de cada pessoa.

A Figura 2 mostra um trecho do XSLT do padrão oai-capes.

Figura 2 - Trecho do XSLT do oai-capes

```
<!--  dcterms:title  -->
<xsl:for-each select=" $dc/*[local-name()='element' and @name='title']/*[local-name()='field' and @name='value'] |
$dc/*[local-name()='element' and @name='title']/*[local-name()='element' and not(@name='alternative')]/*[local-
name()='field' and @name='value'] | $dc/*[local-name()='element' and @name='title']/*[local-name()='element' and
not(@name='alternative')]/*[local-name()='element']/*[local-name()='field' and @name='value'] ">
  <dcterms:title>
    <xsl:value-of select="normalize-space(.)"/>
  </dcterms:title>
</xsl:for-each>
```

Fonte: Elaborado pelos autores (2026).

O modelo de arquivo de transformação (XSLT), que converte os metadados internos do repositório para o formato do padrão oai-capes, é disponibilizado publicamente e deve ser adaptado considerando a planilha de mapeamento dos metadados, explicitada na seção 4.2.1. Essa adaptação requer conhecimentos técnicos em XSLT.

A validação com amostras reais de registros XOAI permitiu identificar e corrigir problemas que não seriam visíveis por inspeção estática da folha de estilo, como a ausência de emissão de campos preenchidos opcionalmente e a geração de elementos vazios quando determinados campos não estão presentes no registro.

4.2.3 entregáveis do padrão oai-capes disponíveis para a comunidade

A partir da concepção e da criação do Padrão oai-capes, foram criados três entregáveis para que a comunidade acadêmica os utilize na implementação desse padrão em repositórios institucionais.

O primeiro documento é o "Guia de Configuração para a Instalação do oai-capes", elaborado para apoiar as instituições na adoção do padrão oai-capes. Tal guia estrutura-se em seções que cobrem a arquitetura do módulo OAI-PMH e o papel dos *crosswalks* XSLT; a estrutura interna do modelo xoai e sua correspondência com os campos *Dublin Core* qualificado; o mapeamento de metadados específico para instituições; os padrões do Perfil de Aplicação da CAPES e as decisões técnicas do *crosswalk*; e os procedimentos de validação e depuração.

Além do guia, foram disponibilizadas uma folha de estilo XSLT completa, comentada e pronta para uso, e uma planilha de mapeamento de metadados que pode ser adaptada pelos RI das IES a partir de seus próprios esquemas de metadados. A documentação foi redigida em linguagem acessível a profissionais de Tecnologia da Informação, Biblioteconomia e Ciência da Informação, e inclui exemplos práticos e glossários de termos técnicos. Tais documentos estão disponíveis publicamente no site do GoPG da CAPES[9].

4.2.4 Aplicação do padrão oai-capes

Após as definições conceituais e aplicadas do Padrão oai-capes, foi realizada uma etapa de validação. Para tal, duas instituições que compõem o GoPG participaram do projeto-piloto. Destaca-se que, para cada instituição, houve uma estratégia. Para a primeira, a própria equipe que criou o padrão oai-capes elaborou o arquivo XSLT para o *crosswalk* de metadados. A segunda instituição utilizou o "Guia de Configuração para a Instalação do oai-capes" para a elaboração do *crosswalk* de metadados e a configuração do padrão oai-capes.

No primeiro caso, em que a instituição realizou apenas a instalação do arquivo XSLT desenvolvido pela equipe da RICA|PG, o objetivo consistiu em validar todo o processo do Padrão oai-capes, além de identificar eventuais problemas durante o processo de instalação. Assim, a equipe da RICA|PG realizou a análise dos metadados do repositório institucional e elaborou o mapeamento e o *crosswalk* com a conversão dos metadados do repositório para

[9] Disponível em: https://sucupira.capes.gov.br/guia-de-interoperabilidade/etapas-implementacao/etapa3-interoperabilidade-tecnica/protocolo-repositorios/oai-pmh-capes

aqueles esperados pelo Padrão oai-capes. Coube à equipe de TI da IES apenas configurar o Padrão oai-capes diretamente no servidor do RI. Esse processo foi essencial para a identificação de problemas e desafios encontrados durante os processos e para a elaboração do "Guia de Configuração do oai-capes".

Após esse procedimento e com o "Guia de Configuração do oai-capes" já validado, uma segunda instituição participante do GoPG realizou o processo de instalação do padrão oai-capes. Essa segunda validação teve como objetivo verificar se o acesso ao guia, à planilha e ao XSLT de modelo seria suficiente para a equipe do RI implantar o padrão oai-capes. Nessa instituição, o mapeamento foi realizado pela própria equipe do repositório institucional, e a equipe de tecnologia da informação criou o arquivo XSLT. Da mesma forma, o processo de instalação apoiou a validação do Padrão oai-capes e dos arquivos que apoiam sua implantação, o que foi igualmente importante para o aprimoramento do processo. Vale ressaltar que, em ambas as instituições, o padrão oai-capes foi implantado antes de seu lançamento para a comunidade, demonstrando a validação do padrão desenvolvido.

5 CONSIDERAÇÕES FINAIS

A proposição do Padrão oai-capes para a integração de dados entre repositórios institucionais e a Plataforma Sucupira mostra-se como uma resposta à questão de pesquisa levantada, sendo uma alternativa viável para a conversão de metadados de RI que utilizam perfis de aplicação próprios para a estrutura de dados da Plataforma Sucupira.

O desenvolvimento do Padrão oai-capes criou entregáveis que auxiliam as equipes dos repositórios institucionais vinculados às IES na implantação do padrão e na disponibilização de dados para o processo de avaliação da pós-graduação pela CAPES. Isso auxilia a integração de dados de dissertações e teses com precisão e completude pela CAPES. Foram disponibilizados três entregáveis: a planilha para o mapeamento dos metadados entre os RI e a Plataforma Sucupira, o arquivo *Extensible Stylesheet Language Transformations* (XSLT), que materializa o *crosswalk* de metadados, e o Guia de Configuração do oai-capes.

Vale destacar, ainda, que o padrão oai-capes foi lançado oficialmente pela CAPES, estando validado e em uso por duas instituições que estão integradas ao GoPG.

Como limitação, aponta-se a necessidade de mais IES implantarem o padrão oai-capes, o que permitirá que o Guia de Configuração seja aprimorado a partir do retorno das demais instituições. Além disso, aponta-se como limitação a vinculação técnica ao DSpace, ao xoai e

ao XSLT 1.0. Como perspectiva futura, o padrão oai-capes será expandido para atender outros tipos de produções intelectuais bibliográficas, técnicas e tecnológicas, artísticas e culturais, o que ampliará a sua abrangência para o contexto apresentado.

Por fim, aponta-se que o trabalho contribui para a área de Informação e Tecnologia ao sistematizar conhecimento científico, técnico e operacional sobre integração de dados de RI, bem como ao evidenciar a forma como decisões de implementação afetam diretamente a qualidade dos dados coletados pela CAPES e, em última instância, a visibilidade da produção científica das IES brasileiras.

REFERÊNCIAS

BRASIL. Coordenação de Aperfeiçoamento de Pessoal de Nível Superior. **Instrução Normativa nº 1, de 23 de abril de 2026**. Brasília: CAPES, 2026. Disponível em: https://cad.capes.gov.br/ato-administrativo-detalhar?idAtoAdmElastic=20426. Acesso em: 13 set. 2026.

BRASIL. Coordenação de Aperfeiçoamento de Pessoal de Nível Superior. **Portaria nº 158, de 17 de agosto de 2023**. Brasília: CAPES, 2023. Disponível em: https://cad.capes.gov.br/ato-administrativo-detalhar?idAtoAdmElastic=12742. Acesso em: 13 set. 2026.

CHAN, Lois Mai; ZENG, Marcia Lei. Metadata interoperability and standardization: a study of methodology, part I. **D-Lib Magazine**, [*s. l.*], v. 12, n. 6, jun. 2006. DOI: https://doi.org/10.1045/june2006-chan. Acesso em: 8 set. 2026.

HASLHOFER, Bernhard; KLAS, Wolfgang. A survey of techniques for achieving metadata interoperability. **ACM Computing Surveys**, New York, v. 42, n. 2, p. 1-37, 2010. DOI: https://doi.org/10.1145/1667062.1667064. Acesso em: 8 set. 2026.

HEERY, Rachel; PATEL, Manjula. Application Profiles: mixing and matching metadata schemas. **Ariadne**, [*s. l.*], n. 25, 2000. Disponível em: https://www.ariadne.ac.uk/issue/25/app-profiles/. Acesso em: 1 mar. 2026.

MÉNDEZ RODRÍGUEZ, Eva María. **Metadatos y recuperación de información**: estándares, problemas y aplicabilidad en bibliotecas digitales. Gijón, Asturias: Ediciones Trea, 2002.

NILSSON, Mikael; BAKER, Thomas; JOHNSTON, Pete. **The Singapore Framework for Dublin Core Application Profiles**. [*S. l.*]: DCMI, 2008. Disponível em: https://www.dublincore.org/specifications/dublin-core/singapore-framework/. Acesso em: 8 set. 2026.

OPEN ARCHIVES INITIATIVE. **The Open Archives Initiative Protocol for Metadata Harvesting**. Version 2.0. [*S. l.*]: Open Archives Initiative, 2002. Disponível em: https://www.openarchives.org/OAI/openarchivesprotocol.html. Acesso em: 10 abr. 2026.

ST. PIERRE, Margaret; LAPLANT, William P. **Issues in crosswalking content metadata**

standards. Bethesda: NISO, 1998. Disponível em: https://www.niso.org/publications/crosswalking-content-metadata-standards. Acesso em: 10 abr. 2026.

VAN DE SOMPEL, Herbert; LAGOZE, Carl. The Santa Fe Convention of the Open Archives Initiative. **D-Lib Magazine**, [*s. l.*], v. 6, n. 2, fev. 2000. DOI: https://doi.org/10.1045/february2000-vandesompel-oai. Acesso em: 10 abr. 2026.

VAN DE SOMPEL, Herbert; NELSON, Michael L. Reminiscing about 15 years of interoperability efforts. **D-Lib Magazine**, [*s. l.*], v. 21, n. 11/12, nov./dez. 2015. DOI: https://doi.org/10.1045/november2015-vandesompel. Acesso em: 10 set. 2026.

WOODLEY, Mary S. Crosswalks, metadata harvesting, federated searching, metasearching: using metadata to connect users and information. *In*: BACA, Murtha (ed.). **Introduction to metadata**. 3. ed. Los Angeles: Getty Research Institute, 2016.